COVID-19 Misinformation on Twitter: Multilingual Analysis

Raj Ratn Pranesh[a,*], Mehrdad Farokhenajd[b], Ambesh Shekhar[a] and Genoveva Vargas-Solar[c]

[a]Birla Institute of Technology, Mesra, India
[b]Univ. Grenoble Alpes, CNRS, LIG, Grenoble, France
[c]Univ. Grenoble Alpes, CNRS, LIG-LAFMIA, Grenoble, France

ARTICLE INFO

Keywords:
Misinformation
Multilingual Analysis
Twitter
COVID-19

ABSTRACT

In the current scenario of the coronavirus disease pandemic (COVID-19), the Internet has become an important source of health information for users worldwide. During pandemic situations, myths, sensationalism, rumors and misinformation, generated intentionally or unintentionally, spread rapidly through social networks. Twitter is one of this popular social networks people use to share COVID-19 related news, information, and thoughts that reflect their perception and opinion about the pandemic. Analysis of tweets for identifying misinformation can generate valuable insight to evaluate the quality and readability of online information about the COVID-19.

This paper presents a multilingual COVID-19 related tweet analysis method, CMTA, that uses BERT, a deep learning model for multilingual tweet misinformation detection and classification. CMTA extracts features from multilingual textual data, which is then categorized into specific information classes. Classification is done by a Dense-CNN model trained on tweets manually annotated into information classes (i.e., 'false', 'partly false', 'misleading'). The paper assesses CMTA experimenting an analysis of multilingual tweets from February to June, showing the distribution type of information spread across different languages.

1. Introduction

Since late 2019, the coronavirus disease COVID-19 has been spreading globally affecting over 216 countries [27]. COVID-19 has created a massive impact on multiple sectors including countries economy, government bodies, private companies, media houses and most importantly, affecting the mental and physical health of human beings by tempering their daily routine activities [38, 12].

COVID-19 also has made us realize how well the world is interconnected through the Internet. Social media is a significant conduit where people share their response, thoughts, news, information related to COVID-19, with one in three people in the world engaging in social media, and two-thirds of those on the Internet using it [28]. Studies have shown that many people connect to the Internet and social media platforms such as Twitter, Facebook, Whatsapp, Instagram and Reddit every day and utilizing it for getting information/news through them [25] [18]. Twitter users, in particular, are known for sharing and consuming news: 59% of Twitter users describe it as good or extremely good for sharing preventive health information [40].

However, social media is also rife with health misinformation. For users with nonmedical education, it is difficult to judge the reliability of health information on the Internet. Accurate and credible dissemination of right information about the virus causing a pandemic could help in controlling the spread of the virus and associated anxiety in the population [36]. Information and misinformation on social media can influence public opinion and behaviour with intense consequences, positively or negatively manipulating the perspective of those who consume it [3, 22]. In the Munich security conference held on February 15, 2020, the general director of WHO commented, We are not just fighting an epidemic; we are fighting an infodemic [41]. It is clear that there is no way to prevent the spread of COVID-19, but it is important to verify the information on the internet to prevent the panic and misinformation associated with the disease. The greatest problem of Internet health information is finding valid and reliable information [10].

In the case of COVID-19, misinformation comes in many forms such as 'COVID-19 is a biological weapon created either by the US (to destroy Chinese) or China (to destroy Americans), COVID-19 is the likely by-product of Chinese foods, like bat soups among other foods, unverified home remedies like vitamin C, cow urine, turmeric etc., self-detection test by holding breath. At its worst, misinformation of this sort may cause people to turn to ineffective (and potentially directly harmful) remedies, as well as to either overreact (e.g., by hoarding goods) or, more dangerously, underreact (e.g., by deliberately engaging in risky behaviour and inadvertently spreading the virus) [3, 29]. Unfortunately, the fake news spread faster than the virus [14].

An online social platform such as Twitter provides particularly fertile ground for the spread of misinformation [13]. Twitter provides direct access to an unprecedented amount of content and may amplify rumours and questionable information [5]. With such a huge amount of human-generated information being exchanged every day, it has attracted Natural Language Processing (NLP) researchers to explore, analyze, and generate valuable insights about people response to COVID-19. People response is analyzed with respect to sentiments and misinformation and malicious information

*Corresponding author
raj.ratn18@gmail.com (.R.R. Pranesh); Mehrdad.Farokhnejad@univ-grenoble-alpes.fr (M. Farokhenajd); ambesh.sinha@gmail.com (.A. Shekhar); genoveva.vargas@imag.fr (G. Vargas-Solar)
ORCID(s):

detection.

This paper proposes CMTA, a multilingual tweet analysis and information (misinformation) detection method for understanding both the negative and positive sides of social media during COVID-19 pandemic. CMTA uses Multilingual BERT, trained on 104 multiple languages to derive features from tweets and 1D convolution for finding the correlation between data of hidden states. It also uses a dense layer for linear transformation on contextual embeddings to provide inferential points. Our work helps in providing better results in finding the proximity of being fake. We used manually annotated tweet misinformation for training two separate deep neural network model training a classifier model for detecting and identifying the type of information (disinformation) present in tweets.

For experimenting with our method, we used trained models for a systematic analysis of COVID-19 related tweets collected from February to June 2020. The analysis of tweets is done based on the distribution of the type of information present in tweets concerning the language used for writing a tweet. We investigated the presence of false information spread throughout Tweeter by classifying the tweets in three classes: 'false', 'partly false' and 'misleading'. We have provided illustrative statistical representation of our findings and detailed discussion about the insights discovered in our survey. The **motivation** for designing a multilingual method lies behind the need of analyzing not just monolingual tweets but also multilingual tweets by building a single deep learning framework that would be able to understand tweets in multiple languages.

The remainder of the paper is organised as follows. Section 2 introduces disinformation background and the notions that guide our study. It then introduces works that have addressed misinformation detection about COVID-19 on social media datasets. Section 3 describes the datasets and the data preparation process adopted in our study. Section 4 describes the method CMTA that we propose, including the general approach behind the method. Section 5 describes the experiment setting that we used for applying CMTA. It also discusses the results of the study. Sections 6 discuss results obtained by applying CMTA to perform multilingual analysis of tweets. Finally, Section 8 concludes the paper and discusses future work.

2. Background and Related Work

In this section, we describe the background of misinformation and why it is important to address the problem of misinformation. We have also summarized the past work done in the field of COVID-19 misinformation detection and analysis.

2.1. Misinformation

According to the Oxford English Dictionary, both misinformation[1] and disinformation[2] are either wrong or misleading information. Misinformation involves information that is accidentally false and is shared without intent to cause harm, whereas disinformation refers to false information knowingly being created and shared to cause harm [17].

Literature provides more fine-grained definitions of various examples of misinformation on social media, including rumor, fake news, hoax, satire, propaganda, and even conspiracy [34, 26, 6], such that: (1) fake news is the presentation of fake or misleading claims as news, where the claims are misleading deliberately [15], (2) hoax is deliberately fabricated falsehood, with the intention to deceive a certain group of the population [15], (3) rumor is unverified but relevant information that arise in contexts such as danger or potential threat, that helps people make sense and manage risk [9].

In reality, claims are not always completely false or true but can have piece of false information or inaccurate information [33]. Such claims are coined partially false [32].

In this work, we are interested in a general notion of misinformation and do not differentiate between misinformation and disinformation as it is virtually impossible to computationally determine ones intention.

2.2. Looking for COVID-19 misinformation in social media

The COVID-19 pandemic has resulted in immense growth in studies that have been published to investigate the various types of misinformation arising during the COVID crisis [2, 8, 37, 22]. They either investigate a very small subset of claims [37] or they manually annotate a small subset of Twitter data [22]. In [2] authors analyse different types of sources looking for COVID-19 misinformation. They show that the majority appear on social media outlets. Pennycook et al. [29] introduced an attention-based account of misinformation and observed that people tend to believe false claims about COVID-19 and share false information when they do not think critically about the accuracy and veracity of the information. Kouzy et al. [22] annotated about 600 messages containing hashtags about COVID-19, they observed that about one-fourth of messages contain some form of misinformation and about 17% contain some unverifiable information. With such misinformation overload, any decision making procedure based on misinformation has high likelihood of severely impacting people's health [20]. [19] examined the global spread of information related to crucial disinformation stories and fake news URLs during the early stages of the global pandemic on Twitter. Their study shows that news agencies, government officials, and individual news reporters do send messages that spread widely, and so play critical roles. However, the most influential tweets are those posted by regular users, some of whom are bots. Tweets mentioning fake news URLs and misinformation stories are more likely to be spread by regular users than the news or government accounts.

[35] focused on sentiment analysis and topic modeling and designed a dashboard to track misinformation on Twitter regarding the COVID-19 pandemic. The dashboard provides an analysis of topics, sentiments, and trends, assessed

[1] https://www.oed.com/view/Entry/119699?redirectedFrom=misinformation
[2] https://www.oed.com/view/Entry/54579?redirectedFrom=disinformation

from Twitter posts; along with identified false, misleading and clickbait information spreading on social media, related to COVID-19. Cinelli et al.

[37] are monitoring the flow of (mis)information flow across 2.7M tweets, and correlating it with infection rates to find that misinformation and myths are discussed, but at lower volume than other conversations.They observed that a meaningful spatio-temporal relationship exists between information flow and new cases of COVID-19, and while discussions about myths and links to poor quality information exist, their presence is less dominant than other crisis specific themes.

In [16] proposed a first example of causal inference approach to discover and quantify causal relationships between pandemic characteristics (e.g. number of infections and deaths) and Twitter activity as well as public sentiment. They observed that their proposed method could successfully capture the epidemiological domain knowledge and identify variables that affect public attention and perception.

An infodemic observatory analysing digital responses in online social media to COVID-19 has been created by CoMuNe lab at Fondazione Bruno Kessler (FBK) institute in Italy, and is available online [3]. The observatory uses Twitter data to quantify collective sentiment, social bot pollution, and news reliability and displays this visually.

Based on the geo-tagged dataset from the US on a state and county level, [11] analyzed tweets to study the daily tweeting patterns in different states. First, they could detect differences in temporal tweeting patterns and found that most state pairs have a strong linear correlation and hourly tweeting behaviors show that people tweeting more about COVID-19 during working hours. In addition, they used facial emojis to track the different types of public sentiment during pandemic including an event specific subtask reporting negative sentiment when the 100th and 1000th death was announced and positive when the lockdown measures were eased in the states.

[23] explored the discourse around the COVID-19 pandemic and government policies being implemented. They used Twitter data from different countries in multiple languages and identify common responses to the pandemic and how these responses differ across time using text mining. Moreover, they presented insights as to how information and misinformation were transmitted via Twitter.Similarly, [31] use text mining on Twitter data to show the epidemiological impact of COVID-19 on press publications in Bogota, Colombia. Intuitively, they find that the number of tweets is positively correlated with the number of infected people in the city.

Most of the works described above focus on analysing tweets related to single language such as English. In our work we have designed a single model leveraging multilingual BERT for the analysis of tweets in multiple languages. Furthermore, we used a large data set to train and analyze the tweets. Our aim is to provide a system that will be restricted to any language for analysing social media data.

[3] https://covid19obs.fbk.eu/

3. Data preparation

This section discusses the steps involved in the collection of COVID-19 related tweets. For training our misinformation detection deep learning model, we have extracted annotated misinformation data from multiple publicly available open databases. We also collected a very large number of multilingual tweets consisting of over 2 million tweets belonging to eight different languages.

3.1. Training Dataset

In order to train and test our misinformation detection model, we collected the training data from an online fact-checker website called Poynter [30]. Poynter have a specific COVID-19 related misinformation detection program named 'CoronaVirusFacts/DatosCoronaVirus Alliance Database[4]'. This database contains thousands of labelled social media information such as news, posts, claims, articles about COVID-19 which were manually verified and annotated by human volunteers(fact-checkers) from all around the globe. The database gathers all the misinformation related to topics such as COVID-19 cure, detection, the effect on animals, foods, travel, government policies, crime, lockdown.

The misinformation dataset was available in 2 languages- 'English' and 'Spanish'. Since we were training a multilingual BERT model, we crawled through the content of all 2 websites using Beautifulsoup[5], a Python library for scraping information from web pages. We scrape 8471 English language false news/information belonging to nine major classes namely, 'False', 'Partially false', 'Misleading', 'No evidence', 'Four Pinocchios', 'Incorrect', 'Three Pinocchios', 'Two Pinocchios' and 'Mostly False'. For each article we gathered the article's title, it's content and the fact checker's misinformation-type label. Similarly, from the Spanish[6] databases we collected 531 misinformation articles respectively. The collected data contains the misinformation published on social media platforms such as Facebook, Twitter, What'sapp, YouTube and were mostly related to political-biased news, scientifically dubious information and conspiracy theories, misleading news and rumors about COVID-19. We also used one more human annotated fact-checked tweet dataset [1] available at the public repository[7]. The dataset contained true and false labelled tweets in English and Arabic language. We used only false labelled tweets consisting of 500 English. We compiled (table 2) a total of 9,502 micro-articles distributed across 9 misinformation classes.

Defining misinformation classes: The collected data was unevenly distributed across 9 classes. We put the classes such as 'No evidence', 'Four Pinocchios[8]', 'Incorrect', 'Three Pinocchios[9]', 'Two Pinocchios[10]' and 'Mostly False' under the minority group because of having very few labels. On the

[4] https://www.poynter.org/covid-19-poynter-resources/
[5] Python module is available at https://pypi.org/project/beautifulsoup4/
[6] https://chequeado.com/latamcoronavirus/
[7] https://github.com/firojalam/COVID-19-tweets-for-check-worthiness
[8] 90%-95% changes of it being false
[9] 70%-75% changes of it being false
[10] 50%-55% changes of it being false

Our Rating	IFCN(Poynter) Rating	Misinformation	Explanation
False	False	The border between France and Belgium will be closed.	French and Belgian authorities denied it.
	Four pinocchios	Trumps effort to blame Obama for sluggish coronavirus testing.	There was no Obama rule, just draft guidance that never took effect and was withdrawn before President Trump took office.
	Inaccurate	Elisa Granato, the first volunteer in the first Europe human tria of a COVID-19 vaccine, has died.	Elisa Granato, the first volunteer in the first Europe human trial of a COVID-19 vaccine, has died.
Partially False	Partially False	Media shows a Florida beach full of people while its empty.	The different videos were not shot at the same time. The beaches are empty when they are closed.
	Two Pinocchios	The bill for a coronavirus test in the US is $3.000	The CDC is not making people pay the test by now.
	Partly False	Salty and sour foods cause the body of the COVID-19 virus" to explode and dissolve.	Consuming fruit juices or gargling with warm water and salt does not protect or kill COVID-19, the World Health Organization Philippines told VERA Files.
Misleading	Misleading	A clip from Mexico depicts the dumping of coronavirus patients corpses into the sea.	Misbar's investigation of the video revealed that it does not depict the dumping of coronavirus patients corpses in Mexico, but rather paratroopers landing from a Russian MI 26 helicopter.
	No Evidence	Media uses photos of puppets on patient stretchers to scare the public.	There is no evidence that any media outlet used this photo for their reporting about COVID-19. Its origin is unclear, maybe it was shot in Mexico and shows a medical training session.
	Mostly False	Coronavirus does not affect people with O+ blood type.	The post claiming coronavirus does not affect people with O+ blood type is misleading.

Table 1
Misinformation Dataset

Classes	Number of tweets
False [30] (English)	2,869
Partially False (English)	2,765
Misleading (English)	2,837
False (Spanish)	191
Partially False (Spanish)	161
Misleading (Spanish)	179
False [1] (English)	500
Total	9,502

Table 2
Collected Misinformation Dataset

other hand, labels like 'False', 'Partially false' and 'Misleading' comprises the majority group as most of the collected articles belongs to this group. In order to structure and distribute the dataset uniformly for training our model, we reformed the dataset by merging the minority group labels into the majority group labels. The classes ('Four Pinocchios' and 'Incorrect') that correspond to completely false information were merged together into the 'False' class. 'Three Pinocchios' and 'Two Pinocchios' were merged together into 'Partially false' class. 'No evidence' and 'Mostly False' were put together with the 'Misleading' class.

Table 1 gives a clear understanding of our training dataset and showcase some misinformation articles present in our training dataset. Column 1 shows the reformed label assigned by us, column 2 shows the original label assigned by the fact-checker, column 3 gives a misinformation example associated with the label present in column 2, and column 4 provides a reasoning given by the fact-checker behind assigning a particular label (column 2) to the misinformation (column 3). For example, if we would look at the entry number '3' in the table 1, the misinformation is about the adverse effect of 5G radiation over the COVID-19 patients. This was labeled 'Incorrect' by the fact-checker. After analysing the fact-checker rating and the explanation given, we labelled it as 'False' misinformation. Entry number '5' talks about the COVID-19 test cost. The explanation given by fact-checker is valid as it is not sure if there is any fee in USA for COVID-19 test or not. So because of the lack of evidence and uncertainty we labelled it as 'Partially false'. Entry number '7' in the table talks about a video showing COVID-19 corpus dumping in the sea. Based on the explanation, the video was coupled with the wrong information to mislead the audience.

Language	ISO	Number of tweets
English	en	1,472,448
Spanish	es	353,294
Indonesian	in	80,764
French	fr	71,722
Japanese	ja	71,418
Thai	th	36,824
Hindi	hi	27,320
German	de	23,316
Sum		2137106

Table 3
Language-wise Dataset Distribution

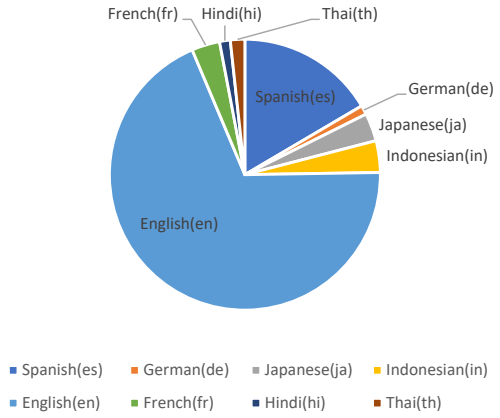

Figure 1: Language-wise Dataset Distribution Pie chart.

So it was labelled as 'Misleading' misinformation.

3.2. Inference Dataset

Once we finished training our multilingual tweet misinformation detection model we aimed to use it for predicting and analysing the misinformation spread across all over the social media platforms in multiple languages. In order to do so, we collected around 2,137,106 multilingual tweets consisting of tweets belonging to eight major languages, namely- 'English', 'Spanish', 'Indonesian', 'French', 'Japanese', 'Thai', 'Hindi' and 'German'. We used an ongoing dataset of tweets IDs associated with the novel coronavirus COVID-19 [4]. Started on January 28, 2020, the current version of dataset contains 212,978,935 tweets divided into groups based on their publishing month. The dataset was collected using multilingual COVID-19 related keywords and contains tweets in more than 30 languages. We used tweepy[11] which is a Python module for accessing twitter API. For our analysis we decided to retrieve the tweets using the tweet IDs of the tweets published in past 5 months (February, March, April, May and June). Table 3 shows the total number of tweets collected by us and figure 1 shows their distribution across eight different language.

[11]Python module is available at http://www.tweepy.org

4. The CMTA Method

In this section, we have given a detailed sequential overview of CMTA method design. We utilize the self-attention mechanism of the BERT for text feature extraction, CNN for exploiting local correlation of the data and dense layer for linear transformation.

In CMTA, the BERT model we are adapting is a multilingual based bidirectional transformer, which is trained on 104 multiple languages. Its architecture resembles the BERT-base model with 12 encoding layers and 110M parameters and resolves the normalization issues faced in different languages. The tokenizer from Multilingual BERT helps in tokenizing inputs of different languages by generating embeddings for the network. BERT generally gives two outputs, one pooled output also called contextual embeddings, and another hidden-states of each layer. We use both of these for further processing.

We use the dense layer or fully connected layer for linear transformation of the data by matrix-vector multiplication with Rectified Linear unit as activation, the dense layer performs a sequence of translation, rotation, and scaling based on the value of kernels and bias.

To handle the sequence data, 1D convolution proves to be a better option. Since Conv1D can handle the spatial dimension and are known for really fast computations, they are the best efficient alternatives to traditional recurrent neural networks. Just like 2D Convolutions, we can also perform operations like padding, striding, or dilation in our architecture. In this way, Conv1D can use for hidden state values for the correlation of data.

4.1. Analytics Pipeline and Methodology

Figure 2 shows the phases of the analytics pipeline of CMTA with their internal processes consisting of four phases: tokenizing, text features extraction, linear transformation, local correlation of data, and classification after the concatenation.

Tokenizing for Multilingual-BERT Before everything, we need our data compatible with the network, so we will convert our textual data to numerical data using the tokenization method. Since this tokenization has to be done for our BERT model, we will be using BERT's tokenizer for multilingual data. The length of the string that should be tokenized will be limited to 512, any string greater than this much of tokens will be truncated, otherwise padded from the right. We tokenize our string to two vectors: input vectors, and segment vectors. These two vectors have a dimensionality of 128 for each sequence and contains relevant id for each token.

Multilingual BERT: The feature Extractor In this phase, we will be utilizing the attention mechanism of BERT on the text. Since our text is vectorized into numerical data, these vectors will be able to extract contextual features using attention mechanisms from encoder-decoder of the layers of the BERT. These values are then sent to the next encoder by a feed-forward network where Softmax is applied to normalize

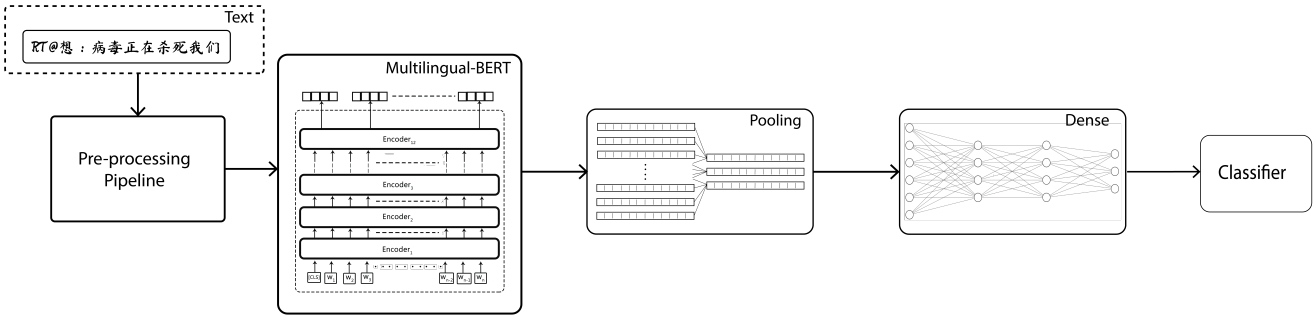

Figure 2: A detailed structure of CMTA architecture.

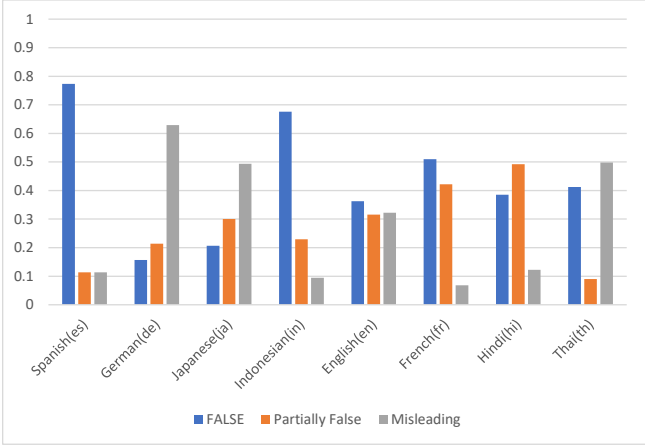

Figure 3: Language-wise Disinformation Distribution.

the output. A vector of a dimension of 768 for each token is generated by the first encoder, which moves through every layer of the BERT network for calculation till the last layer.

Linear Transformation with the dense layer This phase comprises of Pooling and Linear Transformation of pooled data with training the architecture. The extracted features from BERT is processed in this phase. We perform two different types of pooling on the extracted features and reshape the output into a linear vector, so that Dense layer can perform linear transformation like scaling, translation and various linear algebraic operation on the data, and with the help of the back-propagation, we normalize the value of the gradients so that this phase could be shaped perfectly to provide inferential output. To avoid vanishing gradient we use LeakyReLU and dropout layers. Depending upon the kernel value and bias value we get our final processed output for further operation.

Classification In the end, a linear layer of size 3 is connected to the model in the end for classification. This classification layer outputs a Softmax value of vector, depending on the output, the index of the highest value in the vector represents the label for the given sequence.

5. Experiment

5.1. Dataset Proprocessing

In data preprocessing, we performed cleaning and structuring of the training and inference dataset. The collected dataset contained lots of unnecessary noises and components such as emojis, symbols, numeric values, hyperlinks to websites and username mentions which were needed to be removed. Since our dataset was multilingual, we had to be very careful while preprocessing as we did not wanted to lose any valuable information. We used simple regular expressions to remove URLs, special characters or symbols, blank rows, re-tweets, user mentions but we did not removed the hashtags from the data. As hashtags might contain useful information. For example in the sentence- 'Wear mask to protect yourself from #COVID-19 #corona', only '#' symbol was removed during the preprocessing(e.g. 'Wear mask to protect yourself from COVID19 corona'). We removed stop words using NLTK[12], a Python library for natural language processing. NLTK supports multiple languages except few languages such as Hindi and Thai in our case. For preprocessing Hindi dataset we used CLTK(Classical Language Toolkit) [13] which supports Hindi stop words. For removing Thai stop words from Thai tweets, we used PyThaiNLP [39].

[12] https://www.nltk.org/
[13] https://docs.cltk.org/en/latest/index.html

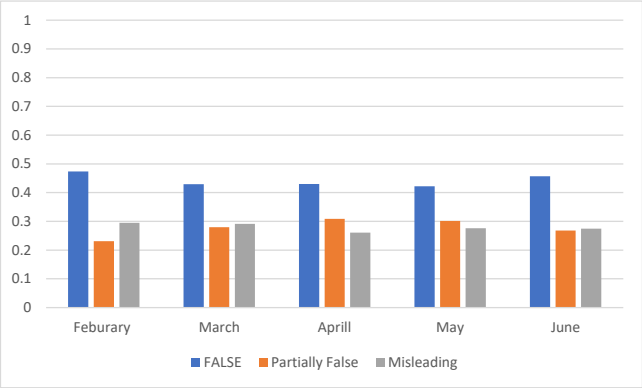

Figure 4: Month-wise Disinformation Distribution.

Test Data	Actual Label	Prediction	Accuracy(✓/✗)
Dr. Megha Vyas from Pune, India died due to COVID-19 while treating COVID patients.	False	False	✓
El plátano bloquea la entrada celular del COVID-19	False	False	✓
Asymptomatic people are very rarely contagious, said the WHO.	Partially False	Partially False	✓
Patanjali Coronil drops can help cure coronavirus.	Misleading	Misleading	✓
El medicamento contra piojos sirve como tratamiento contra Covid-19.	Misleading	False	✗

Table 4
Misinformation data examples along with model's prediction and actual label

The emojis were removed using their unicodes. For training our model we divided the dataset into training, validation and testing dataset in the ratio of 80%/10%10% respectively. The final count for train, validation and test dataset was 7,602, 950, 950.

5.2. Model Setup and Training

Training Setting We fine-tuned the Sequence Classifier from HuggingFace based on the parameters as specified in [7]. Thus, we set a batch size of 32, learning rate 1e-4, with Adam Weight Decay as the optimizer. We run the model for training for 10 epochs. Then, we save the model weights of the transformer. These will be helpful for the further training.

Hyperparameters' Setting Table 5 lists every hyperparameter for training and testing our model. All the calculations and selection of hyperparmaters are done based on tests and for the best output from the model. After performing several iterations on distinct sets of hyper-parameters, based on the analysis of the model's performance, we adopted the one showing promising results on our dataset.

Parameters	Value
Pool Size of Average Pooling	8
Pool Size of Max Pooling	8
Dropout Probability	0.36
Number of Dense layers	4
Text Length	128
Batch Size	32
Epochs	10
Optimizer	Adam
Learning Rate	1×10^{-4}

Table 5
Hyper-parameters for training

5.3. Results assessment

This section discusses the performance our multilingual model over the test data. On the test dataset, our model was able to achieve an accuracy(%) of **82.17** and F_1(%) of **82.54**. The precision and recall reported by the model were **82.07** and **82.30** respectively. Table 4 shows model's prediction over few examples from the test dataset along with their actual label. As we shown in the table, the model prediction in case of entry number '1', '2', '3' and '4' our model was able to predict the correct the label. But in case of entry number '5' the label predicted by our model was 'False' whereas the actual label is 'Misleading'. If we would look at the misinformation at the entry number '5' which is a Spanish text- 'El medicamento contra piojos sirve como tratamiento contra Covid-19.' and who's English translation would be-". This misinformation claims about a COVID-19 medicine and since this could be 'false' and 'misleading' misinformation at the same time, our model predicted it as a 'false' misinformation rather than 'misleading'.

Figure 5: Month-wise Disinformation Distribution in Languages.

6. Multilingual Misinformation Analysis

In this section, we provide a detailed analysis misinformation distribution across the multilingual tweets. We used our trained multilingual model to predict and categorize the misinformation type present in tweets. We conducted our sequential misinformation analysis on a collection of over 2 million multilingual tweets. Our survey studied and analyzed the distribution of COVID-19 misinformation across eight major languages, (i.e. 'English', 'Spanish', 'Indonesian', 'French', 'Japanese', 'Thai', 'Hindi' and 'German') for five months (i.e. February, March, April, May and June). Figure 5 shows the month-wise distribution of misinformation types for each language. Table 6 presents a detailed count of misinformation classes across all the languages. In

the figure 4, we could observe that for February, March and June months our model predicted large number of tweets as 'False', followed by 'Misleading' which is second largest and the number of 'Partially false' was the least. For the tweets generated during the month of April and May, our model discovered that the number of 'Partially false' tweets are more than 'Misleading' tweets and 'False' tweets were again in majority. Figure **??** parallelly showcase the overall(all 5 months together) spread of misinformation types across each language. We could clearly see that German tweets have the highest number of 'Misleading' tweets whereas French have the least. Spanish tweets beats other language's tweets by becoming the language with largest source of 'False' misinformation. Germany generated the least number of 'False' tweets. Hindi tweets tends to have the highest number of 'Partially false' tweets whereas Thai have the least of all. Following more specific observation made with respect to the languages:

- English: The misinformation distribution for English data, indicates that there is a majority of **False** tweets during the five months, whereas the distribution of **Misleading** labelled data is slightly less than as compared to **False** labelled data. **Partially False** labelled tweets are moderately distributed, as in month April we can see that there is a greater number with respect to other months.
- Spanish: From the distribution graph, Spanish tweets have greater frequency of **False** labelled tweets, whereas the **Misleading** tweets and **Partially False** tweets shows almost same number of tweet across the five months.
- German: There was a surge of **Misleading** labelled tweets during the month February, and the count remained the same throughout the five months. There was also an increase in **Partially False** tweets in March but it decreased in successive months, leading to minor **False** labelled tweets.
- Japanese: In the graph of language wise-distribution5, it can be seem that on an average throughout the five months, approx 20% of Japanese tweets are labelled **False**, similarly approx 30% of the Japanese tweets are labelled **Partially False**, leading to the majority of 50% data are labelled as **Misleading**. We can also see that there was a huge increase in **Misleading** tweets in March, tweeted in Japanese language.
- Indonesian: In our distribution for Indonesian tweets approximately 10% of tweets are labelled as **Misleading** and in contrary there is a large distribution of **False** labelled tweets. Approximately 34% of the data in Indonesian dialect is labelled as **Partially False** throughout the five months.
- French: Figure5 shows the misinformation distribution across all of the five months in the French tweets. The largest majority of the tweets were classified as **False** misinformation. Among **Partially false** and **Misleading**, the least number of tweets were labelled as **Misleading**.
- Hindi: The frequency of Hindi tweets is low in the dataset used in our experiment. Yet, our model can predict or label Hindi tweets. Tweets in Hindi have low numbers of **Misleading** tweets, whereas the **Partially False** tweets class has a great frequency. **False** labelled tweets are slightly low compared to **Partially False** tweets in this dialect.
- Thai: The distribution of Thai tweets, shows that our model prediction is majorly oriented towards the **Misleading** tweets. The distribution of **Misleading** labelled tweets it the greatest among the labelled classes, in contrast to **Partially False** tweets. **False** labelled tweets are comparatively moderate in this language.

7. CMTA vs Monolingual BERT Models

In this section, we have presented a comparative performance study of various monolingual BERT models with respect to our proposed multilingual CMTA model for the misinformation detection task. We investigated eight monolingual BERT model[14], namely, 'English', 'Spanish', 'French', 'Germann', 'Japanese', 'Hindi' 'Thai[15]' and 'Indonesian'.

Data Processing: We utilized the same 9,502 tweets distributed across 3 misinformation classes for training the monolingual models. Since our dataset was consist of tweets in English and Spanish language; we translated the tweets into eight languages for training each of the eight monolingual model. We used Google Translator API[16] for converting the tweets into a particular language.

Experiment and Result: We experimented the multilingual data with their respective linguistic based BERT models. We set the model training parameters same as the CMTA model, and preprocessed the data as stated previously. Each of the monolingual model was fine-tuned for 10 ephocs with batch size of 32. using the classification dataset of their respective language.

EnglishBERT scored an F1-score of 77.9% on the English tweets, with recall rate of 74.18%. This possible reason could be that it is heavily trained on English Corpus. From huggingface's model library we got SpanishBERT. The model scored an F1-score of 76.2% with recall rate of 72.02% and precision 80.9%. For French tweets we used CamemBERT[24] from huggingface. The CamemBERT scored an F1-score of 76.32%, with recall rate of 71.45% and precision 81.91%. GermanBERT showed a significant results on German-basesd tweets. It had a precision of 80.61% with recall rate of 71.43%, resulting to an F1-score of 75.74%. JapaneseBERT derived from the paper [21], is 79.56% precise on Japanese tweets

[14]Pretrained model available at https://huggingface.co/models
[15]ThaiBERT available at https://github.com/ThAIKeras/bert
[16]Please refer https://cloud.google.com/translate/docs

Lingo	February			March			April		
	Misinformation			Misinformation			Misinformation		
	False	Partially False	Misleading	False	Partially False	Misleading	False	Partially False	Misleading
Spanish	58346	6653	13740	67956	10913	8826	34125	5437	3604
German	517	581	2505	862	1438	3043	584	892	2664
Japanese	1920	3079	5245	448	692	2650	1635	2850	5840
Indonesian	11157	3226	1951	12573	4336	1582	9073	3367	1273
English	88369	62747	76640	92428	96571	105143	77368	74947	63473
French	4464	3472	1155	12024	10270	1670	6650	5300	763
Hindi	500	870	202	756	909	348	2211	2868	705
Thai	1950	1074	2780	6036	736	7678	2263	554	2917

Lingo	May			June		
	Misinformation			Misinformation		
	False	Partially False	Misleading	False	Partially False	Misleading
Spanish	57821	8214	7107	54965	8828	6759
German	1076	1426	4430	616	657	2028
Japanese	8984	12324	18125	1741	2496	3389
Indonesian	12695	4574	1805	9114	3038	1000
English	140494	128326	119391	135172	101896	109483
French	8475	7667	842	4952	3535	483
Hindi	4560	6057	1343	2501	2739	751
Thai	2825	470	1830	2103	486	3122

Table 6
Language-wise predicted misinformation labels of tweets

Metrics / Models	Precision	Recall	F1-score
EnglishBERT	82.03	74.18	77.90
SpanishBERT	80.9	72.02	76.20
CamemBERT	81.91	71.45	76.32
GermanBERT	80.61	71.43	75.74
JapaneseBERT	79.56	65.36	71.76
HindiBERT	79.56	65.68	71.95
ThaiBERT	79.11	66.25	72.11
IndonesianBERT	78.96	65.66	71.69
CMTA	81.52	74.40	77.79

with recall rate of 65.36% and F1-score of 71.76%. HindiBERT model had an F1-score of 71.95%, 79.56% precise with recall rate 65.68%. ThaiBERT scored an F1-score of 72.11%, being 79.11% precise with recall rate 66.25% IndonesianBERT is 78.96% precise, recall rate of 65.66%, resulting to an F1-score of 71.69%..

8. Conclusion and Future Work

In this paper, we presented a BERT based multilingual model for analysing COVID-19 related multilingual tweets. We performed a detailed systematic survey for detecting disinformation spread on the social media platform- Twitter. We were able to detect misinformation distribution across eight major languages and presented a quantified magnitude of misinformation distributed across different languages in last 5 months. We strongly believe that our model can help in filtration of misinformation and factual data present in multiple languages during the pandemic.

In future, we aim at collecting more annotated training data and performing analysis of a larger multilingual dataset to gain deeper understanding. We aim at improving our model's robustness and contextual understanding for better performance in the classification task. Since analysis was done on a limited dataset the results cannot be generalised. We hope that through our work researchers could gain more deeper insights about misinformation spread across major languages and hence utilizing the information in building more reliable social media platform.

References

[1] Alam, F., Shaar, S., Dalvi, F., Sajjad, H., Nikolov, A., Mubarak, H., Martino, G.D.S., Abdelali, A., Durrani, N., Darwish, K., Nakov, P., 2020. Fighting the covid-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society. arXiv:2005.00033.

[2] Brennen, J.S., Simon, F., Howard, P.N., Nielsen, R.K., 2020. Types, sources, and claims of covid-19 misinformation. Reuters Institute 7.

[3] Brindha, M.D., Jayaseelan, R., Kadeswara, S., 2020. Social media reigned by information or misinformation about covid-19: a phenomenological study .

[4] Chen, E., Lerman, K., Ferrara, E., 2020. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. JMIR Public Health and Surveillance 6, e19273.

[5] Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C.M., Brugnoli, E., Schmidt, A.L., Zola, P., Zollo, F., Scala, A., 2020. The covid-19 social media infodemic. arXiv preprint arXiv:2003.05004 .

[6] Cui, L., Lee, D., 2020. Coaid: Covid-19 healthcare misinformation dataset. arXiv preprint arXiv:2006.00885 .

[7] Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 .

[8] Dharawat, A.R., Lourentzou, I., Morales, A., Zhai, C., 2020. Drink bleach or do what now? covid-hera: A dataset for risk-informed health decision making in the presence of covid19 misinformation .

[9] DiFonzo, N., Bordia, P., 2007. Rumor psychology: Social and organizational approaches. American Psychological Association.

[10] Eysenbach, G., Powell, J., Kuss, O., Sa, E.R., 2002. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. Jama 287, 2691–2700.

[11] Feng, Y., Zhou, W., 2020. Is working from home the new norm? an observational study based on a large geo-tagged covid-19 twitter dataset. arXiv preprint arXiv:2006.08581 .

[12] Fernandes, N., 2020. Economic effects of coronavirus outbreak (covid-19) on the world economy. Available at SSRN 3557504 .

[13] Frenkel, S., Alba, D., Zhong, R., 2020. Surge of virus misinformation stumps facebook and twitter. The New York Times .

[14] Gallotti, R., Valle, F., Castaldo, N., Sacco, P., De Domenico, M., 2020. Assessing the risks of" infodemics" in response to covid-19 epidemics. arXiv preprint arXiv:2004.03997 .

[15] Gelfert, A., 2018. Fake news: A definition. Informal Logic 38, 84–117.

[16] Gencoglu, O., Gruber, M., 2020. Causal modeling of twitter activity during covid-19. arXiv preprint arXiv:2005.07952 .

[17] Hernon, P., 1995. Disinformation and misinformation through the internet: Findings of an exploratory study. Government information quarterly 12, 133–139.

[18] Hitlin, P., Olmstead, K., 2018. The science people see on social media. pew research center.

[19] Huang, B., Carley, K.M., 2020. Disinformation and misinformation on twitter during the novel coronavirus outbreak. arXiv preprint arXiv:2006.04278 .

[20] Ingraham, N.E., Tignanelli, C.J., 2020. Fact versus science fiction: fighting coronavirus disease 2019 requires the wisdom to know the difference. Critical Care Explorations 2.

[21] Kikuta, Y., 2019. Bert pretrained model trained on japanese wikipedia articles. https://github.com/yoheikikuta/bert-japanese.

[22] Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M.B., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E.W., Baddour, K., 2020. Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter. Cureus 12.

[23] Lopez, C.E., Vasu, M., Gallemore, C., 2020. Understanding the perception of covid-19 policies by mining a multilanguage twitter dataset. arXiv preprint arXiv:2003.10359 .

[24] Martin, L., Muller, B., Suárez, P.J.O., Dupont, Y., Romary, L., de la Clergerie, É.V., Seddah, D., Sagot, B., 2019. Camembert: a tasty french language model. arXiv preprint arXiv:1911.03894 .

[25] Matsa, K.E., Shearer, E., 2018. News use across social media platforms 2018| pew research center. Journalism and Media .

[26] Molina, M.D., Sundar, S.S., Le, T., Lee, D., 2019. fake news is not simply false information: a concept explication and taxonomy of online content. American Behavioral Scientist , 0002764219878224.

[27] Organization, W.H., et al., 2020. Coronavirus disease 2019 (covid-19): situation report, 188 .

[28] Ortiz-Ospina, E., 2020. The rise of social media. URL: https://ourworldindata.org/rise-of-social-media.

[29] Pennycook, G., McPhetres, J., Zhang, Y., Lu, J.G., Rand, D.G., 2020. Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. Psychological science 31, 770–780.

[30] Poynter Institute, ., 2020. The international fact-checking network. URL: https://www.poynter.org/ifcn/.

[31] Saire, J.E.C., Navarro, R.C., 2020. What is the people posting about symptoms related to coronavirus in bogota, colombia? arXiv preprint arXiv:2003.11159 .

[32] Shahi, G.K., Dirkson, A., Majchrzak, T.A., 2020. An exploratory study of covid-19 misinformation on twitter. arXiv preprint arXiv:2005.05710 .

[33] Shahi, G.K., Nandini, D., 2020. Fakecovid–a multilingual cross-domain fact check news dataset for covid-19. arXiv preprint arXiv:2006.11343 .

[34] Shao, C., Ciampaglia, G.L., Flammini, A., Menczer, F., 2016. Hoaxy: A platform for tracking online misinformation, in: Proceedings of the 25th international conference companion on world wide web, pp. 745–750.

[35] Sharma, K., Seo, S., Meng, C., Rambhatla, S., Liu, Y., 2020. Covid-19 on social media: Analyzing misinformation in twitter conversations. arXiv preprint arXiv:2003.12309 .

[36] Sharma, M., Yadav, K., Yadav, N., Ferdinand, K.C., 2017. Zika virus pandemicanalysis of facebook as a social media health information platform. American journal of infection control 45, 301–302.

[37] Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., Padden, C., Vanarsdall, R., Vraga, E., Wang, Y., 2020. A first look at covid-19 information and misinformation sharing on twitter. arXiv preprint arXiv:2003.13907 .

[38] Torales, J., OHiggins, M., Castaldelli-Maia, J.M., Ventriglio, A., 2020. The outbreak of covid-19 coronavirus and its impact on global mental health. International Journal of Social Psychiatry , 0020764020915212.

[39] Wannaphong Phatthiyaphaibun, Korakot Chaovavanich, C.P.A.S.L.L.P.C., 2016. PyThaiNLP: Thai Natural Language Processing in Python. URL: http://doi.org/10.5281/zenodo.3519354, doi:10.5281/zenodo.3519354.

[40] Wilford, J., Osann, K., Wenzel, L., 2018. Social media use among parents of young childhood cancer survivors. Journal of Oncology Navigation & Survivorship 9.

[41] Zarocostas, J., 2020. How to fight an infodemic. The Lancet 395, 676.